

Redes bayesianas: una herramienta probabilística en los modelos de distribución de especies

R.F. Ropero^{1*}, P. A. Aguilera¹, A. Fernández², R. Rumí²

(1) Departamento de Biología y Geología, Escuela Politécnica Superior y Facultad de Ciencias Experimentales, Universidad de Almería. Carretera Sacramento s/n, La Cañada de San Urbano, 04120. Almería.

(2) Departamento de Matemáticas, Escuela Politécnica Superior y Facultad de Ciencias Experimentales, Universidad de Almería. Carretera Sacramento s/n, La Cañada de San Urbano, 04120. Almería.

* Autor de correspondencia: R.F. Ropero [rosa.ropero@ual.es]

> Recibido el 30 de octubre de 2013, aceptado el 13 de febrero de 2014.

Ropero, R.F., Aguilera, P.A., Fernández, A., Rumí, R. 2014. Redes bayesianas: una herramienta probabilística en los modelos de distribución de especies. *Ecosistemas* 23(1):54-60. Doi.: 10.7818/ECOS.2014.23-1.08

Las redes bayesianas son un modelo probabilístico multivariante que permite trabajar con incertidumbre. En general, su aplicación en los modelos de distribución de especies es escasa y centrada en el tratamiento de datos discretos sin aprovechar toda su potencialidad. En este trabajo se estudian las redes bayesianas como una herramienta para resolver distintos problemas en los modelos de distribución de especies: clasificación, caracterización y regresión. La posibilidad de trabajar con datos discretos y continuos simultáneamente, la variedad de problemas que pueden resolver, y la flexibilidad en la estructura del modelo, convierten a las redes bayesianas en una herramienta apropiada, en los modelos de distribución de especies, en particular y en Macroecología, en general

Palabras clave: caracterización; clasificación; estadística multivariante; modelos gráficos probabilísticos; regresión.

Ropero, R.F., Aguilera, P.A., Fernández, A., Rumí, R. 2014. Bayesian networks: a probabilistic tool for species distribution models. *Ecosistemas* 23(1):54-60. Doi.: 10.7818/ECOS.2014.23-1.08

Bayesian networks are multivariate probabilistic models able to deal with uncertainty. They have been hardly applied in species distribution models, and have mainly focused on discrete variables without taking advantage of their potentiality. In this paper, Bayesian networks are presented as a tool to solve different problems in species distribution models such as classification, characterization and regression. Their ability to deal with discrete and continuous data simultaneously, the variety of problems that can be solved, and the flexibility in the model structure, make them an appropriate tool in species distribution models and Macroecology.

Key words: classification; characterization; probabilistic graphical models; multivariate statistics; regression

Introducción

El desarrollo de los Modelos de Distribución de Especies (MDE) ha estado ligado al avance de los Sistemas de Información Geográfica (SIG) y las técnicas estadísticas (Guisan y Zimmermann 2000, Segurado y Araújo 2004). Algunas de las técnicas aplicadas han sido: los árboles de clasificación (Miller y Franklin 2002, Dzeroski y Drumm 2003, Fukuda et al. 2013), análisis de correspondencias canónicas (Guisan et al. 1999), distintos métodos de regresión (Lehmann et al. 2002a y b, Manel et al. 2001, Li y Wang 2013), modelos específicos como BIOCLIM (Busby 1986, 1991, Aragón et al. 2013), FLORAMAP (Jones y Gladkov 1999, Erre et al. 2009), o DOMAIN (Walker y Cocks 1991, Ortega-Huerta y Peterson 2008) o técnicas de inteligencia artificial como las redes neuronales (Moisen y Frescino 2002, Dedecker et al. 2004, Fukuda et al. 2013).

En general, las técnicas empleadas pretenden relacionar datos de presencia, presencia-ausencia o abundancia de especies con variables ambientales (Elith et al. 2006) con el objetivo principal de obtener mapas de distribución de las especies (Anderson et al. 2003, Guisan y Thuiller 2005). Gráficamente, estas relaciones se pueden representar como una red de interacciones causales sujetas a incertidumbre.

Las redes bayesianas (RBs) son una herramienta estadística aplicada en entornos bajo condiciones de incertidumbre. En una revisión llevada a cabo por Aguilera et al. (2011) se observa que las RBs están siendo muy desarrolladas a nivel matemático, mientras que su aplicación en el campo de la modelización ambiental es escasa. Más del 80 % de los artículos revisados emplean datos de tipo discreto, pese a la capacidad de las RBs para tratar datos continuos e, incluso, híbridos (datos continuos y discretos simultáneamente). Puesto que fueron originariamente propuestas para el tratamiento de variables discretas, existe una amplia bibliografía, metodología y software al respecto (Nyberg et al. 2006). Sin embargo, es muy común que los datos ambientales presenten variables continuas y discretas simultáneamente. La solución más utilizada es la discretización de las variables continuas y su tratamiento como discretas. Este proceso supone una pérdida de información estadística, reduciendo la exactitud y precisión del modelo final (Uusitalo 2007).

El interés por el tratamiento de variables continuas y discretas, dentro del mismo modelo, ha propiciado el desarrollo de nuevas técnicas. El primer modelo descrito fue el modelo Gausiano (Lauritzen 1992, Lauritzen y Jensen 2001), que impone ciertas restricciones: no puede aplicarse si las variables no siguen una distribución nor-

mal, ni en el caso de que variables discretas presenten como padres variables continuas. Posteriormente se han desarrollado una serie de modelos como el "Mixtures of Truncated Exponentials" (MTE) (Moral et al. 2001), el "Mixtures of Polynomials" (Shenoy y West 2011) o el "Mixtures of Truncated Basis Functions" (Langseth et al. 2012). Sólo el MTE ha sido aplicado, hasta el momento, en la modelización ambiental (Aguilera et al. 2010, 2013). Todos estos modelos pueden ser aplicados sea cual sea la naturaleza de los datos (discretos, continuos o híbridos) sin imponer restricciones en la estructura.

De la misma forma, en los MDE, las RBs han sido escasamente aplicadas (Lehmkuhl et al. 2001, Raphael et al. 2001, Steventon et al. 2006, Newton et al. 2007, Pollino et al. 2007a y b, Smith et al. 2007, Steventon y Daust 2009, Aguilera et al. 2010, Grech y Coles 2010, Johnson et al. 2010, Dlamini 2011, Jay et al. 2011, Laws y Kerler 2012, Murray et al. 2012). En general, estos trabajos utilizan datos discretos y construyen el modelo de RB mediante expertos para resolver problemas de clasificación y caracterización. De esta tendencia difiere el trabajo de Aguilera et al. 2010 que aplica modelos de estructuras fijas cuyos parámetros se aprenden a partir de los datos, para evaluar la probabilidad de presencia de tortuga mora en Andalucía, utilizando datos de naturaleza híbrida.

Estos trabajos ponen de manifiesto que las RBs se aplican parcialmente, sin aprovechar toda la potencialidad que ofrecen. En este artículo se pretende analizar su aplicabilidad en los MDE, haciendo énfasis en la diferencia entre caracterización, clasificación y regresión, en función del objetivo del modelo, el tipo de variables disponibles (continuas o discretas) y el proceso de aprendizaje de las diferentes estructuras propuestas en la literatura.

Redes Bayesianas

Las RBs (Pearl 1988, Jensen y Nielsen 2007) son una potente herramienta para modelar sistemas complejos bajo incertidumbre (Getoor et al. 2004, Uusitalo 2007). Se definen como un modelo estadístico multivariante para una serie de variables $\mathbf{X} = \{X_1, \dots, X_n\}$ formado por dos componentes:

Componente cualitativo: grafo dirigido acíclico (Fig. 1) compuesto por una serie de nodos conectados entre sí mediante relaciones direccionadas, donde los ciclos no están permitidos. En el caso de las RB, cada nodo representa una variable en el modelo, y los arcos entre nodos codifican las relaciones de dependencia / independencia entre las mismas.

Por su componente cualitativo, las RBs facilitan la comunicación entre los expertos permitiendo, de una manera sencilla, identificar las relaciones de dependencia o independencia entre las variables, y aquellas variables irrelevantes sin necesidad de ningún cálculo previo. Además, presentan la facilidad de incorporar datos e información de fuentes muy diversas. El conocimiento procedente de fuentes no empíricas, como datos bibliográficos o la opinión de expertos, pueden ser introducidos en la RB en cualquiera de los pasos del proceso de aprendizaje y construcción del modelo (Cain et al. 2003, Castelletti y Soncini-Sessa 2007, Henriksen et al. 2007).

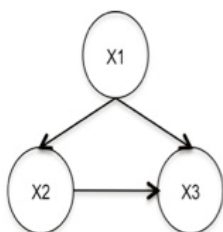


Figura 1. Ejemplo de componente cualitativo de una RB con tres variables (X_1 , X_2 y X_3). La variable X_1 es padre de las variables X_2 y X_3 ; a la vez que X_2 es padre de X_3 .

Figure 1. Example of the qualitative component of a Bayesian network with three variables (X_1 , X_2 y X_3). Variable X_1 is parent of X_2 and X_3 , and X_2 is parent of X_3 .

Componente cuantitativo: distribuciones de probabilidad condicionada para cada variable (X_i) dado sus padres ($pa(X_i)$) en el grafo, expresadas mediante tablas de probabilidad condicionada (TPC) en el caso de variables discretas, y funciones de densidad para las continuas.

El tratamiento de las variables continuas aporta información más detallada (Fig. 2). El valor de probabilidad de las variables discretas viene dado para cada uno de las categorías de la variable (Fig. 2a). Por contra, los resultados de las variables continuas son expresados mediante funciones de densidad a partir de las cuales, es posible calcular cualquier estadístico (media, mediana, cuartiles,...) o probabilidad de interés (Fig. 2b).

A pesar de las ventajas que presenta el uso de variables continuas, apenas se han aplicado en la modelización ambiental (Aguilera et al. 2011). La mayor dificultad de estos modelos y la falta de bibliografía aplicada y software disponible son las principales causas de su escasa utilización (Ropero et al. 2013).

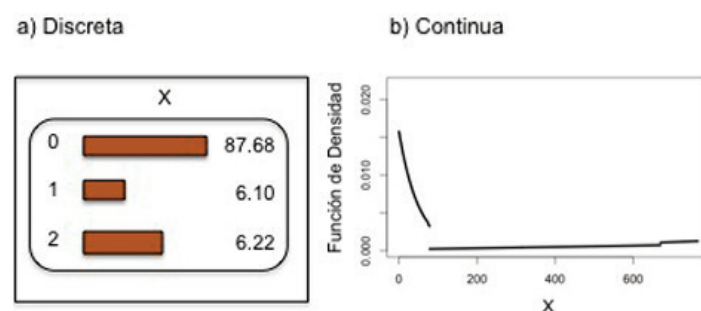


Figura 2. Ejemplos de resultados obtenidos con variables discretas (a) y continuas (b). Los valores de probabilidad de las variables discretas vienen expresados para cada intervalo (0, 1 y 2). En el caso de las continuas, se usan funciones de densidad.

Figure 2. Results obtained from an example with discrete (a) and continuous (b) variables. In the discrete case, a probability value for each interval (0, 1 and 2) is given, whilst a density function is used for continuous variables.

Redes Bayesianas en los Modelos de Distribución de Especies

En los MDE, las RBs pueden aplicarse con tres objetivos principales (Fig. 3): caracterización, clasificación y regresión.

RBs para Caracterización

En este caso se pretende modelizar el sistema de interacciones más o menos complejas entre la especie y su entorno. Una vez construido el modelo, se analiza la presencia y la naturaleza de las relaciones establecidas entre las variables y se plantean escenarios futuros.

Si bien este tipo de RB es interesante, revisando la literatura existente (ver Introducción), no se ha aprovechado hasta ahora toda su potencialidad. En general, se utilizan para construir modelos no demasiado complejos, sin trabajar con grandes matrices de datos o con matrices en las que existan datos faltantes (Raphael et al. 2001, Steventon y Daust 2009, Johnson et al. 2010). La estructura de los modelos han sido construidas únicamente mediante expertos (Nyberg et al. 2006, Castelletti y Soncini-Sessa 2007) en lugar de aprovechar también la potencialidad de los algoritmos de aprendizaje automático, capaces de extraer, a partir de los datos, las relaciones de dependencia/independencia existentes entre las distintas variables. En la mayoría de los casos se emplean datos discretizados, con la consiguiente pérdida de información.

RBs para Clasificación

Resolver un problema de clasificación implica construir un modelo para predecir el valor de una única variable clasificadora de tipo discreto, en función de una serie de variables descriptivas que

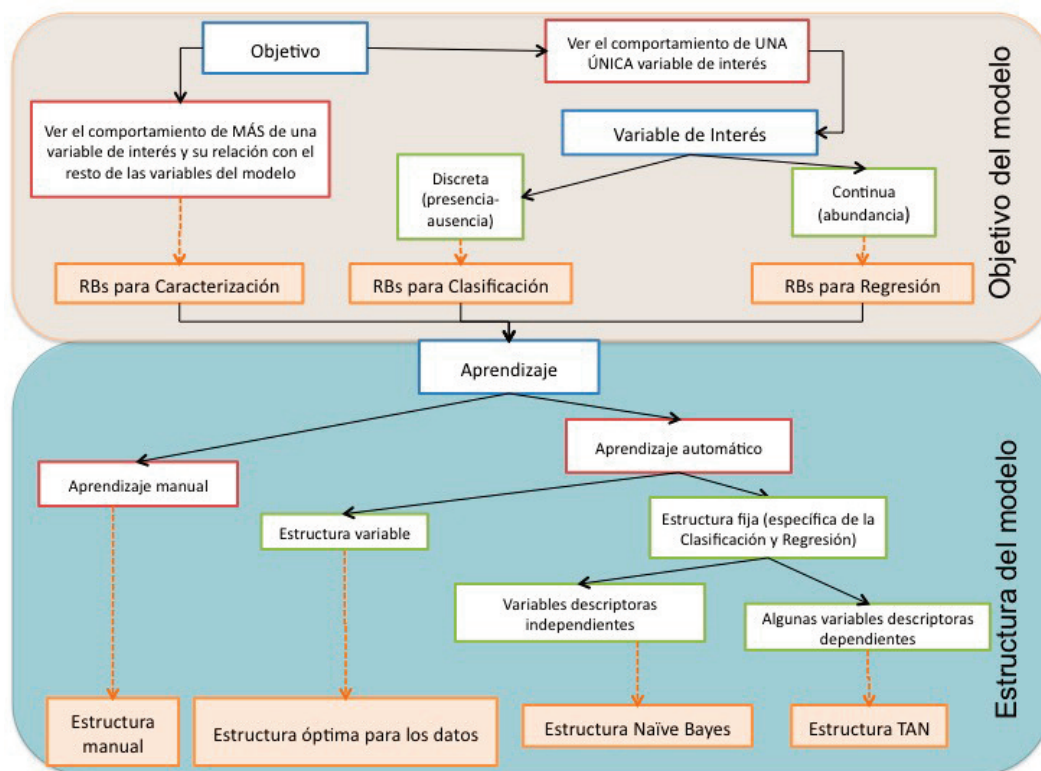


Figura 3. Esquema conceptual de la aplicación de las RBs en los MDE.

Figure 3. Conceptual framework of the application of Bayesian networks in species distribution models.

pueden ser discretas o continuas (Hastie et al. 2009). En este caso, el estudio de relaciones entre las variables descriptoras no es el objetivo prioritario, aunque en algunos problemas su inclusión puede enriquecer el proceso de clasificación.

Estos modelos han sido los más empleados en los MDE (Steven-ton et al. 2006, Smith et al. 2007, Grech y Coles 2010, Dlamini 2011, Jay et al. 2011, Laws y Kesler 2012, Murray et al. 2012), a través de modelos construidos mediante expertos y con datos discretos.

RBs para Regresión

En las RBs los problemas de regresión se resuelven de una forma similar a la clasificación. En este caso, se usan estadísticos para obtener el valor predicho de la variable de interés de tipo continuo, a partir de su función de densidad a posteriori (por ejemplo la esperanza). Las RBs para regresión no se han aplicado en los MDE, aunque podrían suponer un gran avance respecto a los modelos de regresión tradicionales aplicados cuando se tienen datos de abundancia.

Una vez establecido el objetivo del modelo (caracterización, clasificación o regresión), el proceso de aprendizaje consiste en estimar tanto la estructura como los parámetros del modelo final. Las RBs pueden ser aprendidas directamente de los datos, o incluyendo la opinión de expertos en casi todos los pasos del aprendizaje del modelo. La naturaleza de los datos va a determinar el método de aprendizaje más adecuado. Se distinguen tres tipos de aprendizaje (Fig. 3): aprendizaje manual, aprendizaje automático, y una combinación de ambos.

Aprendizaje manual

El aprendizaje manual está siendo muy aplicado en los MDE, en particular, (Pollino et al. 2007a y b, Murray et al. 2012) y en Macroecología, en general (Cain et al. 2003, McDowel et al. 2009). Consiste en la construcción del modelo (su estructura, y/o los valores de las TPC) mediante la opinión de expertos, bibliografía y otras fuentes no empíricas. Este tipo de aprendizaje utiliza datos de tipo discreto y proporciona una estructura de relaciones entre las distintas variables obtenida de forma manual. Si los datos son

continuos, los expertos solo pueden estimar la estructura del modelo, pero no el valor de las funciones de densidad.

Existen diferentes metodologías para la modelización ambiental mediante expertos (Voinov y Bousquet 2010). En el caso concreto de las RBs, Castelletti y Soncini-Sessa (2007) describen el proceso Participatory and Integrated Planning, para construir una RB empleando solo el conocimiento experto. En el mismo sentido, Henriksen et al. (2007), describe el proceso Public Participation. En general, y aunque existen variantes, las metodologías propuestas por estos autores, comienzan con una fase inicial de reconocimiento del problema, las variables involucradas e, incluso, sus diferentes estados. A continuación, se construye la estructura de relaciones entre las distintas variables mediante etapas más o menos complejas de consultas entre los distintos colectivos involucrados (expertos y actores sociales). Por último, se procede a aprender los parámetros del modelo bien mediante los mismos expertos o mediante la inclusión de datos empíricos (aprendizaje mixto).

Aprendizaje automático

Es la aplicación de algoritmos de aprendizaje que, a partir de los datos, infieren la estructura de relaciones más adecuada y los parámetros del modelo. Los algoritmos más utilizados son el PC y K2 (Cooper y Herskovits 1992, Spirtes et al. 1993) para el aprendizaje de la estructura, mientras que los parámetros se obtienen mediante máxima verosimilitud o, mediante el algoritmo EM para el caso de datos faltantes (Dempster et al. 1977).

Al igual que en el aprendizaje manual, se puede aprender automáticamente la estructura, los parámetros del modelo o ambos simultáneamente, dando lugar a dos posibles soluciones (Fig. 4): las estructuras adaptadas a los datos, donde la estructura varía en cada caso en función de los datos con los que se trabaja, y las estructuras fijas, en las que solamente varían los parámetros del modelo para cada conjunto de datos.

Estructuras adaptadas a los datos

Este grupo de estructuras son aprendidas específicamente para cada matriz de datos. Son estructuras variables que pueden

adoptar distintos grados de complejidad según el problema que se pretenda resolver. Pueden aplicarse tanto a RBs para caracterizar, clasificar como para resolver problemas de regresión. Son las estructuras más aplicadas en modelización ambiental, si bien en la literatura no se ha aprovechado toda la capacidad que tienen para el tratamiento de variables continuas y matrices con datos faltantes.

Estructuras fijas

Las estructuras fijas se diseñaron como estructuras específicas para los problemas de clasificación y regresión. Han sido aplicadas escasamente en la modelización ambiental (Park y Stenstrom 2006, 2008, Porwal et al. 2006, Fernandes et al. 2010) y apenas en los MDE (Aguilera et al. 2010, Dlamini 2011). Existen diferentes estructuras fijas: naïve Bayes (Minsky 1963), TAN (Friedman et al. 1997), KdB (Lucas 2002) y AODE (Webb et al. 2005). En este trabajo se van a analizar las estructuras naïve Bayes y TAN por su adaptabilidad y aplicabilidad en los MDE (Aguilera et al. 2010).

El naïve Bayes (NB) (Fig. 5) es una estructura sencilla compuesta por una variable padre (Y), la variable de interés, de la que parten las variables descriptoras ($X_1, \dots, X_n$), que son independientes entre sí, dado Y. Esta fuerte asunción de independencia se compensa por la reducción del número de parámetros a aprender. A pesar de esta aparente sencillez, sus resultados son muy competitivos cuando se enfrenta a otros modelos (Friedman et al. 1997).

Esta estructura se adapta fácilmente a los MDE, entendidos como modelos de clasificación donde el objetivo es determinar el valor de la variable dependiente en función de una serie de variables descriptoras (Fig. 5).

Partiendo de una estructura naïve Bayes, se pueden incorporar algunas relaciones entre las variables ambientales, adoptando una estructura TAN (Fig. 6). En este caso, las variables descriptoras $X_1, \dots, X_n$ forman un árbol de relaciones (Fig. 6a) al que se le añade la variable dependiente, Y (Fig. 6b). Este modelo refleja más fielmente la realidad del sistema puesto que incorpora relaciones de dependencia entre las variables descriptoras, sin embargo, el aumento en la complejidad y el número de parámetros a estimar, hace que estos modelos no tengan por qué ser, necesariamente, más precisos que el modelo naïve Bayes.

Aprendizaje mixto

Consiste en una combinación entre los aprendizajes manual y automático. En la mayoría de los casos, donde se aplica este tipo de aprendizaje (Batchelor y Cain 1999), los expertos determinan la estructura completa del modelo, cuyos parámetros son aprendidos, posteriormente, a partir de los datos. Existen numerosos trabajos que, aunque aprenden la estructura y los parámetros directamente de los datos, emplean el conocimiento y opinión experta para modificar el modelo obtenido (Gambelli y Bruschi 2010). Este último caso se puede plantear como una opción viable en la aplicación de las RBs en los MDE, debido a que, si bien los datos son usados para construir el modelo, los expertos pueden aclarar o remarcar aquellas relaciones que se pueden considerar más importantes o relevantes (McDowell et al. 2009).

Un modelo transversal a todos los mencionados son las redes bayesianas orientadas a objetos (RBOO) (Koller y Pfeffer 1997), que permiten separar un sistema muy complejo en varios subsistemas más sencillos relacionados entre sí. De esta forma, es posible evaluar no solo el comportamiento global del sistema, sino también cada uno de los subsistemas por separado (Molina et al. 2009). Estos modelos no han sido aplicados en los MDE ni en Macroecología pero pueden suponer un gran avance en la modelización de sistemas complejos.

Planteamiento de escenarios

En los MDE es muy interesante conocer el efecto que determinados cambios en las variables ambientales producen en la distri-

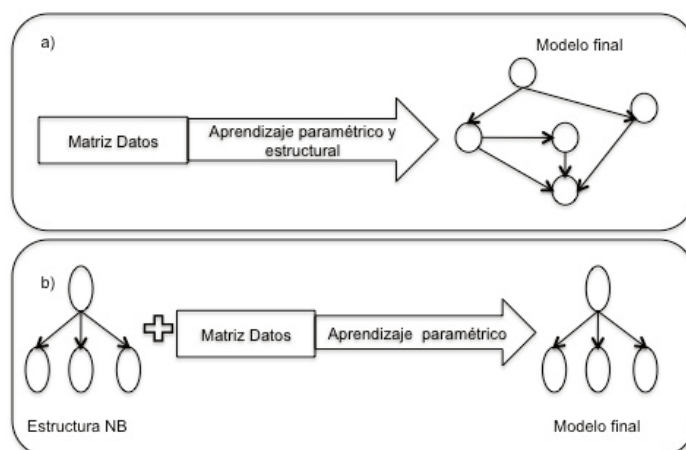


Figura 4. Aprendizaje automático: estructuras adaptadas a los datos (a) y estructuras fijas (b) (NB, naïve Bayes).

Figure 4. Automatic learning: ad-hoc structures (a) and fixed structures (b) (NB, naïve Bayes).

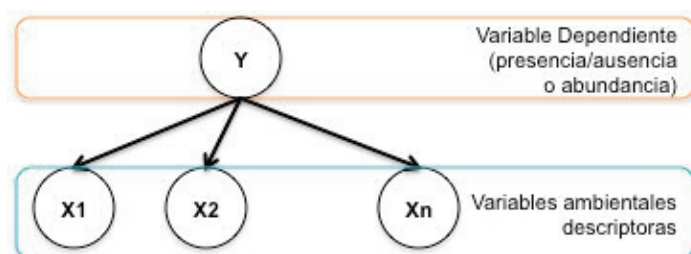


Figura 5. Estructura fija naïve Bayes (NB) para resolver un problema de clasificación (presencia/ausencia) o regresión (abundancia).

Figure 5. Naïve Bayes structure (NB) to solve a classification (presence/absence) or regression (abundance) problem.

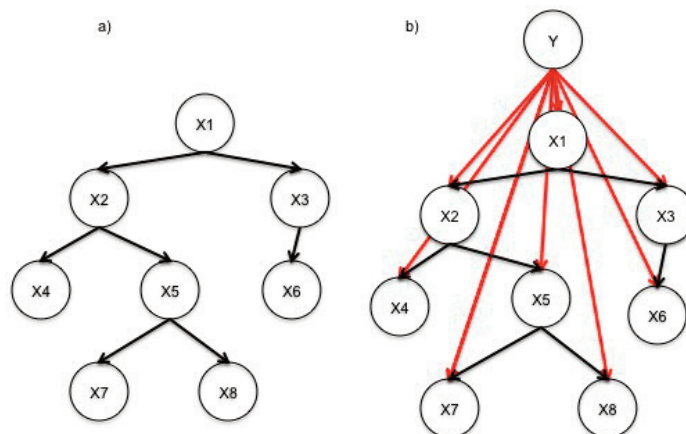


Figura 6. Ejemplo de evolución desde un árbol de relaciones entre las variables descriptoras (a) hasta la estructura TAN al añadir la variable dependiente, Y (b). Modificado de Aguilera et al. (2010).

Figure 6. Extension from a dependence tree (representing the relations among features) (a) into a TAN structure after including the class variable, Y (b). Adapted from Aguilera et al. (2010).

bución de las especies. En el caso de las RBs, una vez construido y validado el modelo (caracterización, clasificación o regresión), es posible incorporar nueva información, conocida como evidencia, en una o más variables con la intención de predecir el comportamiento de otras, lo que se conoce como inferencia probabilística (Fig. 7) (Jensen y Nielsen 2007).

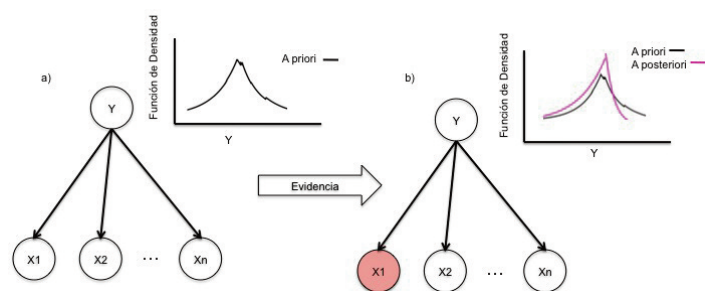


Figura 7. Introducción de nueva información (evidencias) en una RB a priori (a) con el objetivo de determinar el comportamiento a posteriori (b) de una variable de interés.

Figure 7. Including new information (evidence) in a Bayesian network a priori (a) in order to determine the a posteriori behaviour (b) of the variable of interest.

La ventaja que presentan las RBs, frente a otros modelos, es la facilidad con la que permiten plantear escenarios futuros sin tener que introducir información en todas las variables, e independientemente de que las variables sean discretas o continuas. Estas características hacen de las RBs un modelo probabilístico no sólo aplicable en los MDE, sino también en los estudios relacionados con los distintos componentes del cambio global.

Ejemplo de Aplicación

Las RB para clasificación son las más empleadas en los modelos de distribución de especies. En general, estos trabajos emplean datos discretos y construyen el modelo mediante expertos. En este apartado se expone un ejemplo de aplicación obtenido de [Aguilera et al. \(2010\)](#), que aplica un modelo naïve Bayes para evaluar la probabilidad de presencia de tortuga mora (*T. graeca graeca*) en Andalucía, y se compara con una regresión logística en términos de tasa de clasificación.

En ambos casos, el objetivo es relacionar la presencia de tortuga mora (variable discreta) con una serie de variables ambientales descriptoras de naturaleza continua. Se construye un modelo de regresión logística mediante el programa Weka ([Hall et al. 2009](#)), y se compara con el modelo de RB obtenido de [Aguilera et al \(2010\)](#) (ver esta referencia para más información sobre el aprendizaje del modelo) construido mediante el software Elvira ([Elvira-Consortium 2002](#)). Los resultados obtenidos en términos de tasa de clasificación se corresponden con una validación cruzada de 10 hojas.

Entre ambas metodologías, se observa que, si bien son valores similares, se obtiene una mayor tasa de clasificación para el modelo NB ([Tabla 1](#)).

Tabla 1. Valores de tasa de clasificación del modelo naïve Bayes y Regresión logística para la presencia de tortuga mora en Andalucía en el caso de que las variables descriptoras sean discretas o continuas.

Tabla 1. Classification rate of the naïve Bayes and logistic regression models in the study of the spur-thighed tortoise presence in Andalucía with discrete and continuous features.

naïve Bayes	Regresión Logística
97.07 %	96.55 %

Conclusiones y Líneas Futuras

En este trabajo se presentan las RBs como una herramienta aplicable a los MDE capaz de manejar datos discretos, continuos, e incluso ambos en un mismo modelo (modelos híbridos), evitando pérdida de información estadística por la discretización de las va-

riables. A través de las RBs se pueden resolver una variedad de problemas en función del objetivo perseguido (clasificación, caracterización, regresión), ver el comportamiento global del sistema o conocer solo el valor de una única variable. El proceso de aprendizaje del modelo puede llevarse a cabo de diversas formas, a partir de los datos, mediante expertos o una combinación de ambos, adoptando estructuras muy variadas en cuanto a complejidad, desde modelos fijos sencillos como el naïve Bayes, a estructuras adaptadas a los datos con un alto número de relaciones entre las variables.

En la actualidad, el tratamiento de datos continuos a través de distintos modelos como el Gausiano, o los modelos “Mixtures of Truncated Exponentials”, “Mixtures of Polynomials” o “Mixtures of Truncated Basis Functions”, y su aplicación en el campo de la modelización ambiental constituyen una nueva línea de investigación. El análisis de la correlación espacial existente en los datos ambientales, y su tratamiento a través de las RBs es otra línea incipiente de investigación. Por último, las redes bayesianas orientadas a objetos y las redes bayesianas temporales son dos nuevas líneas de desarrollo metodológico apenas aplicadas en la modelización ambiental, ni en los MDE.

Referencias

- Aguilera, P.A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian networks in environmental modelling. *Environmental Modelling and Software* 26: 1376-1388.
- Aguilera, P.A., Fernández, A., Reche, F., Rumí, R. 2010. Hybrid Bayesian network classifiers: Application to species distribution models. *Environmental Modelling and Software* 25: 1630-1639.
- Aguilera, P.A., Fernández, A., Ropero, R.F., Molina, L. 2013. Groundwater quality assessment using data clustering based on hybrid Bayesian networks. *Stochastics Environmental Research Risk Assessment* 27: 435-447.
- Anderson, R.P., Lew, D., Peterson, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162: 211-232.
- Aragon, P., Coca-Abia, M.M., Llorente, V., Lobo, J.M. 2013. Estimation of climatic favourable areas for locust outbreaks in Spain: integrating species' presence records and spatial information on outbreaks. *Journal of Applied Entomology* 137: 610-623.
- Batchelor, C., Cain, J. 1999. Application of belief networks to water management studies. *Agricultural Water Management* 40: 51-57.
- Busby, J.R. 1986. Bioclimate prediction system (BIOCLIM). *User's manual version 2.0*. Australian Biological Resources, Study Leaflet, Canberra, Australia.
- Busby, J.R. 1991. BIOCLIM: A bioclimate analysis and prediction system. En: Margules, C.R., Austin, M.P. (eds.), *Nature conservation: Cost effective biological surveys and data analysis*, pp 64-68. CSIRO, Melbourne, Australia.
- Cain, J.D., Jinapala, K., Makin, I.W., Somaratna, P.G., Ariyaratna, B.R., Perera, L.R. 2003. Participatory decision support for agricultural management. A case study from Sri Lanka. *Agricultural Systems* 76: 457-482.
- Castelletti, A., Soncini-Sessa, R. 2007. Bayesian networks and participatory modelling in water resource management. *Environmental Modelling and Software* 22: 1075-1088.
- Cooper G.F., Herskovits, E.A. 1992. A bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9:309-347.
- Dedecker, A., Goethals, P., Gabriels, W., DePauw, N. 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates communities in the Zwalm river basin (Flanders, Belgium). *Ecological Modelling* 174:161-173.
- Dempster, A.P., Laird, N.M., Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1-38.
- Diamini, W.M. 2011. A data mining approach to predictive vegetation mapping using probabilistic graphical models. *Ecological Informatics* 6: 111-124.
- Dzeroski, S., Drumm, D. 2003. Using regression trees to identify the habitat preference of the sea cucumber (*Holothuria leucospilota*) on Rarotonga, Cook Islands. *Ecological Modelling* 170: 219-226.
- Eliith, J., Graham C.H., Anderson, R.P., Dudik, M., Ferrier, S. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

- Elvira Consortium 2002. Elvira: an environment for probabilistic graphical models. En: *Proceedings of the 1st European Workshop on Probabilistic Graphical Models (PGM'02)*, Cuenca, Spain, November 2002, pp. 222 - 230.
- Erre, P., Chessa, I., Nieddu, G., Jones, P.G. 2009. Diversity and spatial distribution of *Opuntia* spp. In the Mediterranean Basin. *Journal of Arid Environments* 73: 1058-1066.
- Fernández, J.A., Irigoien, X., Goikoetxea, N., Lozano, J.A., Inza, I., Pérez, A., Bode, A. 2010. Fish recruitment prediction, using robust supervised classification methods. *Ecological Modelling* 221: 338-352.
- Friedman, N., Geiger, D., Goldszmidt, M. 1997. Bayesian network classifiers. *Machine Learning* 29: 131-163.
- Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M. 2013. Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environmental Modelling and Software* 47: 1-6.
- Gambelli, D., Bruschi, V. 2010. A Bayesian network to predict the probability of organic farms' exist from the sector: A case study from Marche, Italy. *Computers and Electronics in Agriculture* 71: 22-31.
- Getoor, L., Rhee, J.T., Koller, D., Small, P., 2004. Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine* 30: 233-256.
- Grech, A., Coles, R.G. 2010. An ecosystema-scale predictive model of coastal sea grass distribution. *Aquatic Conservation: Marine and Freshwater Ecosystems* 20: 437-444
- Guisan, A., Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- Guisan, A., Zimmermann, N. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186
- Guisan, A., Weiss, S.B., Weiss, A.D. 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology* 143: 107-122.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. 2009. The WEKA Data Mining Software: An Update; *ACM SIGKDD Explorations Newsletter* 11(1): 10-18.
- Hastie, T., Tibshirani, R., Friedman, J. 2009. *The elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer.
- Henriksen, H.J., Rasmussen, P., Brandt, G., von Bülow, D., Jensen, F.V. 2007. Public participation modelling using Bayesian networks in management of groundwater contamination. *Environmental Modelling and Software* 22: 1101-1113.
- Jay, C.V., Marcot, B.G., Douglas, D.C. 2011. Projected status of the Pacific walrus (*Odobenus rosmarus*) in the twenty-first century. *Polar Biology* 34: 1065-1084.
- Jensen, F.V., Nielsen, T.D. 2007. *Bayesian Networks and Decision Graphs*. Springer.
- Johnson, S., Mergersen, K., Waal, A., Marnewick, K., Cilliers, D., Houser, A.M., Boast, L. 2010. Modelling cheetah relocation success in southern Africa using an Iterative Bayesian Network Development Cycle. *Ecological Modelling* 221: 641-651.
- Jones, P.G., Gladkov, A. 1999. FloraMap: A computer tool for predicting the distribution of plants and other organisms in the Wild. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.
- Koller, D., Pfeffer, A. 1997. Object-Oriented Bayesian Networks. En: *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*, Providence, Rhode Island, USA, 1-3 August, 1997, pp. 302-313.
- Langseth, H., Nielsen, T.D., Rumi, R., Salmerón, A. 2012. Mixtures of Truncated Basis Functions. *International Journal of Approximate Reasoning* 53 (2): 212-227.
- Lauritzen, S.L. 1992. Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association* 87: 1098-1108.
- Lauritzen, S.L., Jensen, F. 2001. Stable local computation with conditional Gaussian distributions. *Statistics and Computing* 11: 191-203.
- Laws, R.J., Kesler, D.C. 2012. A Bayesian network approach for selecting translocation sites for endangered island birds. *Biological Conservation* 155: 178-185.
- Lehmann, A., Overton, J.M., Austin, M. 2002a. Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity and Conservation* 11: 2085-2092.
- Lehmann, A., Overton, J.M., Leathwick, J. 2002b. Grasp: generalized regression analysis and spatial prediction. *Ecological Modelling* 160: 165-183.
- Lehmkuhl, J.F., Kie, J.G., Bender, L.C., Servheen, G., Nyberg, H. 2001. Evaluating the effects of ecosystem management alternatives on elk, mule deer, and White-tailed deer in the interior Columbia River basin, USA. *Forest Ecology and Management* 153: 89-104.
- Li, X.H., Wang, Y. 2013. Applying various algorithms for species distribution modelling. *Integrative Zoology* 8 (2): 124-135.
- Lucas, P. 2002. Restricted Bayesian network structure learning. En: *Proceedings of the 1st European Workshop on Probabilistic Graphical Models (PGM'02)*, Cuenca, Spain, November 2002. pp. 217-232.
- Manel, S., Williams, H.C., Ormerod, S. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921-931.
- McDowell, R.W., Nash, D., George, A., Wang, Q.J., Duncan, R., 2009. Approaches for quantifying and managing diffuse phosphorus exports at the farm/small catchment scale. *Journal of Environmental Quality* 38: 1968-1980.
- Miller, J., Franklin, J. 2002. Modelling distribution off our vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling* 157: 227-247.
- Minsky, M. 1963. Steps towards artificial intelligence. *Computers and Thoughts* 406-450.
- Moisen, G., Frescino, T. 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* 157: 209-225.
- Molina, J.L., Bromley, J., García-Aróstegui, J.L., Sullivan, C., Benavente, J. 2009. Integrated water resources management of overexploited hydrogeological Systems using Object-Oriented Bayesian Networks. *Environmental Modelling and Software* 25: 383-397.
- Moral, S., Rumi, R., Salmerón, A. 2001. Mixtures of Truncated Exponentials in Hybrid Bayesian Networks. En: Benferhat, S., Besnard, P. (eds.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Vol. 2143 of Lecture Notes in Artificial Intelligence, pp. 156-167. Springer-Verlag Berlin Heidelberg, Alemania.
- Murray, J.V., Stokes, K.E., Klinken, R.D. 2012. Predicting the potential distribution of a riparian invasive plants: the effects of changing climate, Floyd regimes and land-use patterns. *Global Change Biology* 18: 1738-1753.
- Newton, A.C., Stewart, G.B., Diaz, A., Golicher, D., Pullin, A.S. 2007. Bayesian Belief Networks as a tool for evidence-based conservation management. *Journal of Nature Conservation* 15: 144-160.
- Nyberg, J.B., Marcot, B.G., Sulyma, R. 2006. Using Bayesian belief networks in adaptive management. *Canadian Journal of Forest Research* 36: 3104-3116
- Ortega-Huerta, M.A., Peterson, A.T. 2008. Modelling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Revista Mexicana de Biodiversidad* 79: 205-216.
- Park, M.H., Stenstrom, M.K. 2006. Using satellite imagery for storm water pollution management with Bayesian Networks. *Water Research* 40: 3429-3438.
- Park, M.H., Stenstrom, M.K. 2008. Classifying environmentally significant urban land uses with satellite imagery. *Journal of Environmental Management* 86: 181-192.
- Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufmann, San Mateo.
- Pollino, C.A., White, A.K., Hart, B.T. 2007a. Examination of conflicts and improved strategies for the management of an endangered Eucalypt species using Bayesian Networks. *Ecological Modelling* 201: 37-59.
- Pollino, C.A., Woodberry, O., Nicholson, A., Korb, K., Hart, B.T. 2007b. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software* 22: 1140-1152.
- Porwal, A., Carranza, E.J.M., Hale, M. 2006. Bayesian network classifiers for mineral potential mapping. *Computers and Geosciences* 32: 1-16.
- Raphael, M.G., Wisdom, M.J., Rowland, M.M., Holthausen, R.S., Wales, B.C., Marcot, B.G., Rich, T.D. 2001. Status and trends of habitats of terrestrial vertebrates in relation to land management in the interior Columbia river basin. *Forest Ecology and Management* 153: 63-88.
- Ropero, R.F., Maldonado, A.D., Aguilera, P.A., Fernández, A., Rumi, R., Salmerón, A. 2013. Discrete vs. Hybrid Bayesian network in Ecological Modelling. En: *Proceedings of the 11th INTECOL Congress, Ecology: Into the next 100 years. 18-23 August 2013, ExCell, London, UK*, pp 80.
- Segurado, P., Araújo, M. 2004. An evaluation of methods for modelling species distribution. *Journal of Biogeography* 31: 1555-1568.
- Shenoy, P.P., West, J.C. 2011. Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning* 52 (5): 641-657.

- Smith, C.S., Howes, A.L., Price, B., McAlpine, C.A. 2007. Using a Bayesian belief network to predict suitable habitat of an endangered mammal-The Julia Creek dunnart (*Sminthopsis douglasi*). *Biological Conservation* 139 (3-4): 333-347.
- Spirtes, P., Glymour, C., Scheines, R. 1993. *Causation, Prediction and Search*. Springer Verlag, Berlin. Alemania.
- Steventon, J.D., Daust, D.K. 2009. Management strategies for a large-scale mountain pine beetle outbreak: Modelling impacts on American Martens. *Forest Ecology and Management* 257: 1976-1985.
- Steventon, J.D., Sutherland, G.D., Arcese, P. 2006. A population-viability-based risk assessment of Marbled Murrelet nesting habitat policy in British Columbia. *Canadian Journal of Forest Research* 36: 3075-3086.
- Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling* 203: 312-318.
- Voinov, A., Bousquet, F. 2010. Modelling with stakeholders. *Environmental Modelling and Software* 25: 1268-1281.
- Walker, P.A., Cocks, K.D. 1991. HABITAT: A procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters* 1: 108-118.
- Webb, G.I., Boughton, J.R., Wang, Z. 2005. Not So Naive Bayes: Aggregating One-Dependence Estimators. *Machine Learning* 58: 5-24.